Improved Nonparametric Estimation of the Sharpe Ratio

* based on joint work with Stefan Sperlich, Jens Perch Nielsen, and Enno Mammen

Michael Scholz
Department of Economics
University of Graz

CFE–CMStatistics
Sevilla, December 2016

years

excess stock returns

1880 1900 1920 1940 1960 1980 2000 2020

−0.6 −0.4 −0.2 0.0 0.2 0.4

Motivation I

Michael Scholz
Objectives of the talk:

- Are equity returns or premiums predictable? Until mid-1980’s: Predictability would contradict the efficient markets paradigm.

- Empirical research in the late 20th century and recent progress in asset pricing theory suggest that excess returns are predictable.

- We take the long-term actuarial view and base our predictions on annual data of the S&P500 from 1872 through 2015 on a one year horizon.

- Our interests:
  - Actuarial models of long-term saving and potential econometric improvements to such models.
  - Market timing/compare assets based on the Sharpe ratio (SR)

\[
SR_t = \frac{\mathbb{E}(Y_t|X_{t-1} = x_{t-1})}{\sqrt{\text{Var}(Y_t|X_{t-1} = x_{t-1})}}
\]
Not many historical years in our records and **data sparsity** is an important issue.

**Bias** might be of **great importance** when predicting yearly data. Classical trade-off of variance and bias depends on the horizon/frequency.

Advocate for non- and semi-parametric methods in financial applications:

- **Powerful data-analytic tools**: Local-linear kernel smoothing and wild bootstrap.
- With **suitable modifications** those techniques can perform well in different economic fields.
- Include **prior knowledge** in the statistical modelling process for **bias reduction** and to **avoid the curse of dimensionality** and other problems.
Overview:

- The prediction framework and the Sharpe ratio
- A measure for the quality of prediction: The validated $R^2$
- Improved smoothing through prior knowledge and estimation of conditional mean/variance function
- Simulation and empirical study
- Outlook and summary
The prediction framework I

- **The excess stock returns:**

\[ S_t = \log\left(\frac{P_t + D_t}{P_{t-1}}\right) - r_{t-1} \]

with **dividends** \( D_t \) paid during period \( t \), **stock price** \( P_t \) at the end of period \( t \), and **short-term interest rate** \( r_t = \log(1 + R_t/100) \) with **discount rate** \( R_t \)

- Consider **one-year-ahead predictions** (\( T = 1 \)), but predictions over the next \( T \) periods are also easily included:

\[ Y_t = S_t + \ldots + S_{t+T-1} = \sum_{i=0}^{T-1} S_{t+i} \]

but this would pose **greater statistical challenges**.
The prediction framework II

- One traditional equation for the value of a stock is
  \[ P_t = \sum_{j=1}^{\infty} (1 + \gamma)^{-j} (1 + g)^{j-1} D_t \]
  with \( \gamma \) discount rate and \( g \) growth of dividend yields.

- Price of stocks depends on quantities such as dividend yield, interest rate, inflation (last two highly correlated with almost any relevant discount rate)

- Covariates \( X_t \) with predictive power: dividend-price ratio, earnings-price ratio, interest rates, . . .

- Consider the model
  \[ Y_t = m(X_{t-1}) + \nu(X_{t-1})^{1/2} \varepsilon_t \]
  where \( \mathbb{E}(\varepsilon_t|X_{t-1}) = 0 \) and \( \text{Var}(\varepsilon_t|X_{t-1}) = 1 \).
A way to examine the performance of an investment by adjusting for its risk (reward-to-variability ratio)

It measures the **excess return** (or **risk premium**) per **unit of deviation** in an investment asset or a trading strategy. **Sharpe (JPM 1994):**

\[
SR_t = \frac{E(Y_t)}{\sqrt{Var(Y_t)}}
\]

Use in finance:

- SR characterizes how well the **return of an asset** compensates the investor for the **risk taken**.
- When comparing two assets vs. a common benchmark, the one with a **higher SR** provides **better return for the same risk** (the same return for a lower risk)
• In practice: **ex-post** SR used with **realized** rather than **expected returns**
  
  \[ SR_{t,\text{simple}} = \frac{\text{mean}(Y_t)}{\text{sd}(Y_t)} \]

• **Easy** to calculate but **depends on length** of observations, includes both systematic and idiosyncratic risk

• **Non-normality** of assets, effect of **covariates, predictions**?

• In our setting:

  \[
  SR_t(x_{t-1}) = \frac{\mathbb{E}(Y_t|X_{t-1} = x_{t-1})}{\sqrt{\text{Var}(Y_t|X_{t-1} = x_{t-1})}} = \frac{m(x_{t-1})}{\sqrt{\nu(x_{t-1})}}
  \]

  and we get a **two-step** estimator for the **Sharpe ratio** as

  \[ \overline{SR_t} = \frac{\hat{m}_t}{\sqrt{\hat{\nu}_t}} \]
Overview:

- The prediction framework and the Sharpe ratio
- A measure for the quality of prediction: The validated $R^2$
- Improved smoothing through prior knowledge and estimation of conditional mean/variance function
- Simulation and empirical study
- Outlook and summary
In detail:

- Consider \( Y_t = \mu + \xi_t \) and \( Y_t = g(X_t) + \zeta_t \)

- \( \mu \) estimated by the mean \( \bar{Y} \) and \( g \) by local-linear kernel regression

- Nielsen and Sperlich (IME 2003) define the validated \( R^2_V \) as

\[
R^2_V = 1 - \frac{\sum_t (Y_t - \hat{g}_{-t})^2}{\sum_t (Y_t - \bar{Y}_{-t})^2},
\]

where the function \( g \) and the simple mean \( \bar{Y} \) are predicted at point \( t \) without the information contained in \( t \)

- A cross-validation criterion to rank different models used for both choice of bandwidth and model selection.
The performance measure II

Properties:

- Replacement of **total variation** and **not explained variation** in usual $R^2$ by its **cross validated** analogs.

- $R^2_V \in (-\infty, 1]$

- Measures how well a given model and estimation principle predicts compared to another (here: to the CV mean)

- If $R^2_V < 0$ we cannot predict better than the mean.

- CV punishes **overfitting**, i.e. pretending a functional relationship that is not really there (leads to $R^2_V < 0$)

- It is a (non-classical) **out-of sample** measure
Overview:

- The prediction framework and the Sharpe ratio

- A measure for the quality of prediction: The validated $R^2$

- Improved smoothing through prior knowledge and estimation of conditional mean/variance function

- Simulation and empirical study

- Outlook and summary
Basic idea: **Combined estimator** Glad (ScanJStat 1998)

Nonparametric estimator **multiplicatively** guided by, for example, parametric model

\[ g(x) = g_\theta(x) \cdot \frac{g(x)}{g_\theta(x)} \]

**Essential fact:**

- Prior captures characteristics of **shape** of \( g(x) \)
- **Correction factor** \( g(x)/g_\theta(x) \) is **less variable** than original function \( g(x) \)
- Nonparametric estimator gives **better results** with **less bias**

**Include** **prior information** in analysis coming from

- Good economic model
- (Simple) empirical data analysis or statistical modelling (**shape**, **constructed variables**, **different frequencies**)
• **Local Problem**: Prior crosses x-axis
  - More robust estimates with suitable **truncation**: Clipping absolute value below $\frac{1}{10}$ and above 10
  - **Shift** by a distance $c$ so that new prior strictly greater than zero and does not intersect the x-axis

• **Dimension reduction**: Use possible **overlapping** covariates $x_1, x_2$ for prior and correction factor
  \[
g(x_1) = (g_\theta(x_2) + c) \cdot \frac{g(x_1)}{g_\theta(x_2) + c}
\]
Scholz et al. (IME 2015) applied the **combined estimator** to annual American data.

A bootstrap test on the true functional form of the conditional expected returns **confirms predictability**.

Including prior knowledge shows **notable improvements** in the prediction of excess stock returns compared to linear and fully nonparametric models.

We will use their **best model** as starting point in our empirical analysis:

- Prior: linear model with **risk-free** rate as predictive variable
- Correction factor: fully nonparametric with **earnings by price** and **long term interest**
- Gives $R^2_V = 20.9$ (compared to a $R^2_V = 14.5$ for a fully nonparametric model without prior)
Four main approaches proposed in the literature: **direct method**, **residual based**, **likelihood-based**, and **difference-sequence method**.

**Direct method** is based on

\[
\text{Var}(Y_t|X_t = x_t) = \mathbb{E}(Y^2_t|X_t = x_t) - \mathbb{E}(Y_t|X_t = x_t)^2
\]

where both parts are **separately** estimated. Result is **not nonnegative** and **not fully adaptive** to the mean fct. Härdle and Tsybakov (JoE 1997) with \( X_t = Y_{t-1} \).

**Residual based** methods consist of two stages:

1. estimate \( \hat{m} \) and squared residuals \( \hat{u}^2_t = (Y_t - \hat{m}(X_t))^2 \)
2. estimate \( \hat{v} \) from \( \hat{u}^2_t = \nu(X_t) + \varepsilon_t \)
Different variants of residual based method (mostly) for 2nd step:

- Fan and Yao (Biometrika 1998) apply \textbf{loc-lin} in both stages. Result is \textbf{not nonnegative} but \textbf{asymp. fully adaptive} to the unknown mean fct.

- Ziegelmann (ET 2002) proposes the \textbf{local exponential estimator} to ensure \textbf{nonnegativity}

\[
\sum_t \left( \hat{u}_t^2 - \psi\{\alpha + \beta(X_t - x)\} \right)^2 K_h(X_t - x) \Rightarrow \text{Min}_{\alpha,\beta}
\]

- Mishra, Su, and Ullah (JBES 2010) propose the use of the \textbf{combined estimator} with a parametric guide. They ignore \textbf{bias} reduction in 1st step.

- Xu and Phillips (JBES 2011) use a \textbf{re-weighted local constant} estimator (maximize the empirical likelihood s.t. a bias-reducing moment restriction).
Yu and Jones (JASA 2004) use estimators based on a **localized normal likelihood** (standard *loc-lin* for estimating the mean $m$ and *loc log-lin* for variance $\nu$)

\[- \sum_t \left\{ \frac{(Y_t - m(X_t))^2}{\nu(X_t)} + \log(\nu(X_t)) \right\} K_h(X_t - x)\]

Wang et al. (AnnStat 2008)

- analyze the **effect of the mean on variance fct. estimation**
- compare the performance of the residual-based estimators to a **first-order-difference-based (FOD) estimator**: *loc-lin* on

\[D_t^2 = \frac{(Y_t - Y_{t+1})^2}{2}\]
Residual-based estimators use an **optimal** estimator for mean fct. in

\[
\hat{\nu}(x) = \sum_t w_t(x)(Y_t - \hat{m}(x_t))^2,
\]

works well if in \( \hat{m} \) the **bias is negligible**. Bias **cannot be further reduced** in 2nd stage. (FOD: crude estimator \( \hat{m}(x_t) = Y_{t+1} \))

Our strategy: Use **combined estimator with simple linear prior** in both stages

Reasons:

- FOD was not convincingly performing in **small samples**
- We know that the mean fct. is rather **smooth**
- But **bias reduction** is key due to **sparsity**
- We cannot compare FOD and residual-based results in terms of \( R^2_V \)
- Maybe FOD together with combined estimator?
Overview:

- The prediction framework and the Sharpe ratio
- A measure for the quality of prediction: The validated $R^2$
- Improved smoothing through prior knowledge and estimation of conditional mean/variance function
- Simulation and empirical study
- Outlook and summary
Simulation Study I

- Which procedure gives reasonable results for estimation of Sharpe ratio?

- Simulate mean and variance function on $x \sim U[0, 1]$ as

  $$m(x) = 2 + \sin(a\pi x), \quad \nu(x) = x^2 - x + 0.75 \quad \text{and} \quad sr(x) = m(x)/\nu(x)^{1/2}$$

- **Best model** in terms of **cross-validated mean square error** (0.020) uses **combined estimator** for mean with a poly3 prior and **combined estimator** for variance with an exponential prior

- **Poorest model** in terms of cross-validated mean square error (0.054) uses **combined estimator** for mean with a exponential prior and **FOD estimator** for variance without prior
averages of best and poorest model over 300 independent estimates \((n = 1000)\)
**Annual American Data**: Updated and revised version of Robert Shiller’s dataset - Chapter 26 in Market volatility (1989)


<table>
<thead>
<tr>
<th></th>
<th>Max</th>
<th>Min</th>
<th>Mean</th>
<th>Sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excess Stock Returns</td>
<td>0.44</td>
<td>-0.62</td>
<td>0.04</td>
<td>0.18</td>
</tr>
<tr>
<td>Dividend by Price</td>
<td>0.09</td>
<td>0.01</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>Earnings by Price</td>
<td>0.17</td>
<td>0.02</td>
<td>0.08</td>
<td>0.03</td>
</tr>
<tr>
<td>Short-term Interest Rate</td>
<td>17.63</td>
<td>0.19</td>
<td>4.61</td>
<td>2.84</td>
</tr>
<tr>
<td>Long-term Interest Rate</td>
<td>14.59</td>
<td>1.91</td>
<td>4.60</td>
<td>2.26</td>
</tr>
<tr>
<td>Inflation</td>
<td>0.17</td>
<td>-0.19</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>Spread</td>
<td>3.27</td>
<td>-5.06</td>
<td>-0.01</td>
<td>1.58</td>
</tr>
</tbody>
</table>
Estimation of the mean fct.:

\[ m(x_1) = (m_\theta(x_2) + c) \cdot \frac{m(x_1)}{m_\theta(x_2) + c} \]

**Tabelle:** Predictive power (in percent)

<table>
<thead>
<tr>
<th></th>
<th>prior</th>
<th>no prior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>d</td>
</tr>
<tr>
<td>corr.</td>
<td>e</td>
<td>8.5</td>
</tr>
<tr>
<td>fac-</td>
<td>e, L</td>
<td>9.8</td>
</tr>
<tr>
<td>tor</td>
<td>e, r</td>
<td>10.9</td>
</tr>
</tbody>
</table>

Our choice: linear model with **risk-free** as prior and fully nonparametric with **earnings by price** and **long-term interest rate** for correction factor.
Empirical Study V

- **estimation of the variance fct.**

\[
\nu(x_1) = \left( \nu_\theta(x_2) + c \right) \cdot \frac{\nu(x_1)}{\nu_\theta(x_2) + c}
\]

**Tabelle:** Predictive power (in percent)

<table>
<thead>
<tr>
<th>factor</th>
<th>prior in mean&amp;var</th>
<th>prior in mean</th>
<th>no prior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( R^2_V ) prior</td>
<td>( R^2_V )</td>
<td>( R^2_V )</td>
</tr>
<tr>
<td>corr.</td>
<td>7.4 L</td>
<td>1.7</td>
<td>0.5</td>
</tr>
<tr>
<td>fac-</td>
<td>5.9 L</td>
<td>1.1</td>
<td>-0.5</td>
</tr>
<tr>
<td>tor</td>
<td>10.7 L, inf</td>
<td>5.4</td>
<td>-1.2</td>
</tr>
</tbody>
</table>

- In all models **long-term interest** or **long-term interest** and **inflations** as prior
- Best model without any prior: \( R^2_V = 1.8 \) with **spread**
Estimator of **Sharpe ratio** evaluated at mean values for other covariates than **inflation** with confidence intervals based on **wild bootstrap**
Overview:

- The prediction framework and the Sharpe ratio
- A measure for the quality of prediction: The validated $R^2$
- Improved smoothing through prior knowledge and estimation of conditional mean/variance function
- Simulation and empirical study
- Outlook and summary
We propose an estimator for the Sharpe ratio as a two stage estimator of conditional mean and variance function using a **combined estimator with parametric priors**.

Include **prior knowledge** in the statistical modelling process. Improve this way smoothing due to **bias reduction**.

We provide an **ex-ante measure of asset performance** incorporating the risk taken.

Could be used for **market timing**. Out-of-sample performance?

SR appropriate measure during **declining markets**?
Thank you for your attention!